SEQUENCE PROTOCOL

GENERAL INFORMATION

APPLICANTS: 1.   Outside lecturer
                               Dr. Christoph von Eichel-Streiber
                               Bingerweg 15
                               55444 Schweppenhausen
                               Federal Republic of Germany
                               Tel.: 06724/33 98
                               Fax : 06724/33 98

                         2.   Prof. Dr. Trinad Chakraborty
                               Seltersweg 85
                               35390 Gießen
                               Federal Republic of Germany
                               Tel.: 0641/76 536
                               Fax : 0641/99 41 259

DESCRIPTION OF THE INVENTION:

    A TGC method for inducting targeted somatic
    transgenesis

NUMBER OF SEQUENCES: 2

POSTAL ADDRESS: Patent attorneys
                         Dr. Rainer A. Keil
                         Ludwig R. Schaafhausen
                         Nanno M. Lenz
                         Dr. K.-H. Meyer-Dulheuer
                         Eysseneckstraße 31
                         D-60322 Frankfurt am Main

COMPUTER-READABLE VERSION

DATA CARRIER: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS DOS
SOFTWARE: Word Perfect 6.0

Information on Sequence ID No. 1:

Length: 1260 base pairs
Type: Nucleic acid and amino acid sequences
                      derived from it
Strand form: single strand
Topology: linear
Origin: Listeria monocytogenes strain EGD
                      Serotype 1/2a
Feature: Sequence of the dapE gene, which is
                      one of the key enzymes needed for
                      synthesis of diaminopimelic acid. The
                      amino acid sequence is highly
                      homologous to N-succinyl-L-
                      diaminopimelic acid desuccinylase
                      (dapE) from e.g. Escherichia coli,
                      Bacillus subtilis, Lactobacillus
                      spp., Mycobacterium tuberculosis.

Amino acid sequence: 318 amino acids
Nucleotide sequence: 1260 nucleotides

```
1     TGCCTTTATA GAGAACGGGA AAACATAGAG TGGAATTCAT AGAAAGAGGG

51    CGTGAAATAT GGACCAACAA AAAAGATTC AAATTTTAAA GGACTTGGTA

101   AATATTGATT CGACTAATGG GCATGAAGAA CAAGTTGCGA ACTATTTGCA

151   AAAGTTGTTA GCTGAACATG GTATTGAGTC CGAAAAGGTA CAATACGACC

201   TAGACAGAGC TAGCCTAGTA AGCGAAATTG GTTCCAGTAA CGA GAA GGT T
                                                    R   E   G

251   TG GCA TTT TCA GGG CAT ATG GAT GTA GTT GAT GCG GGT GAT GTA TCT AAG
       L   A   F   S   G   H   M   D   V   V   D   A   G   D   V   S   K -

301   TGG AAG TTC CCA CCT TTT GAA GCG ACA GAG CAT GAA GGG AAA CTA TAC GG
       W   K   F   P   P   F   E   A   T   E   H   E   G   K   L   Y   G -

351   A CGC GGC GCA ACG GAT ATG AAG TCA GGT CTA GCG GCG ATG GTT ATT GCA A
```

```
                R   G   A   T   D   M   K   S   G   L   A   A   M   V   I   A -
     401   TG ATT GAA CTT CAT GAA GAA AAA CAA AAA CTA AAC GGC AAG ATC AGA TTA
                M   I   E   L   H   E   E   K   Q   K   L   N   G   K   I   R   L -
     451   TTA GCA ACA GTT GGG GAA GAG ATC GGT GAA CTT GGA GCA GAA CAA CTA AC
            L   A   T   V   G   E   E   I   G   E   L   G   A   E   Q   L   T -
     501   A CAA AAA GGT TAC GCA GAT GAT TTA CAT GGT TTA ATC ATC GGC GAA CCG A
                Q   K   G   Y   A   D   D   L   H   G   L   I   I   G   E   P -
     551   GT GGA CAC AGA ATC GTT TAT GCG CAT AAA GGT TCC ATT AAT TAT CCC GTT
             S   G   H   R   I   V   Y   A   H   K   G   S   I   N   Y   P   V -
     601   AAA TCC ACT GGT AAA AAT GCC CAT AGT TCG ATG CCG GAA TCT GGT GTG AA
              K   S   T   G   K   N   A   H   S   S   M   P   E   S   G   V   N -
     651   T GCG ATT GAT AAC TTG CTG CTA TTT TAT AAT GAA GTA GAA AAA TTC GTG A
                A   I   D   N   L   L   L   F   Y   N   E   V   E   K   F   V -
     701   AA TCA GTT GAT GCT ACT AAC GAA ATA TTA GGC GAT TTT ATT CAT AAT GTC
             K   S   V   D   A   T   N   E   I   L   G   D   F   I   H   N   V
     751   ACC GTA ATT GAT GGT GGA AAT CAA GTC AAT AGT ATC CCT GAA AAA GCA CA
            T   V   I   D   G   G   N   Q   V   N   S   I   P   E   K   A -
     801   A CTG CAA GGG AAT ATT CGC TCG ATT CCA GAA ATG GAT AAT GAA ACA GTG A
              L   Q   G   N   I   R   S   I   P   E   M   D   N   E   T   V -
     851   AA CAA GTG CTA GTG AAG ATT ATC AAT AAG TTA AAC AAA CAG GAA AAT GTG
             K   Q   V   L   V   K   I   I   N   K   L   N   K   Q   E   N   V -
     901   AAT CTG GAA TTA ATA TTT GAT TAT GAT AAA CAA CCA GTA TTT AGT GAT AA
              N   L   E   L   I   F   D   Y   D   K   Q   P   V   F   S   D   K -
     951   A AAT TCG GAT TTA GTC CAC ATT GCT AAG AGC GTA GCA AGC GAC ATT GTC
                N   S   D   L   V   H   I   A   K   S   V   A   S   D   I   V
    1001   AAA GAA GAA ATC CCA TTA CTC GGT ATT TCC GGA ACA ACC GAT GCA GCA GA
              K   E   E   I   P   L   L   G   I   S   G   T   T   D   A   A   E -
    1051   A TTT ACC AAA GCT AAG AAA GAG TTC CCA GTG ATT ATT TTT GGA CCA GGA A
                F   T   K   A   K   K   E   F   P   V   I   I   F   G   G   G -
    1101   AC GAA ACC CCT CAC CAA GTA AAC GAA AAT GTT TCT ATA GGA AAT TAT TTG
                N   E   T   P   H   Q   V   N   E   N   V   S   I   G   N   Y   L -
    1151   GAG ATG GTA GAT GTT TAC AAA CGG ATT GCC ACC GAG TTT TTA TCT TGA TGA
                E   M   V   D   V   Y   K   R   I   A   T   E   F   L   S   STOP
    1201   AACTTTAACT TTACTTATTT CCCGATATAA AATAAGTAAT TAATAGAAGT
    1251   CTAGTATTTG 1260
```

Information on Sequence ID No. 2:

Length:	1337 base pairs
Type:	Nucleic acid and amino acid sequences derived from it
Strand form:	single strand
Topology:	linear
Origin:	Listeria monocytogenes strain EGD 1/2a
Feature:	Sequence of the "cold shock protein" cspl; this protein is essential for the viability of Listeria at low temperatures.

Amino acid sequence: 66 amino acids
Nucleotide sequence: 1337 nucleotides

```
  1    GAGGCAAGTG GACTAATCAT AAAGTTTTTG GCGATGCAAC TGCGTTTTG
 51    GCAGGAGATG CTTTACTAAC GCTCGCTTTT TCTATTTTAG CTGAAGACGA
101    TAATTTATCT TTTGAGACAC GCATTGCTTT GATTAACCAA ATTAGTTTTA
151    GTAGCGGTGC AGAAGGAATG GTTGGTGGTC AACTTGCAGA CTTGGAAGCG
201    GAAAACAAAC AAGTGACGCT AGAAGAGTTA TCATCCATTC ATGCACGAAA
251    AACGGGTGAA TTATTAATTT ATGCTGTAAC CTCTGCAGCA AAAATTGCGG
301    AAGCTGATCC AGAACAAACG AAACGCTTAC GAATTTTTGC AGAGAATATT
351    GGGATTGGAT TTCAAATTAG CGACGATATT TTAGATGTAA TTGGTGATGA
401    AACGAAAATG GGTAAAAAGA CAGGGGCCGA CGCTTTTCTG AATAAAAGTA
451    CCTATCCCGG ATTACTCACG CTTGATGGGG CAAAAAGGGC ATTAAATGAG
501    CATGTTACGA TTGCAAAGTC AGCGCTTTGA GGGCATGATT TCGATGATGA
551    AATTCTCTTG AAACTTGCTG ATTTAATCGC ACTTAGAGAA AATTAATCAT
601    AATTATCTAG TAATTTCAAA ATTTTTTCAC ATATATAATT CAAATTGATT
651    TGCTTTTCCT AAAATACCGT GTTATACTAA TGTAAGATTA TTTTTGTGGG
701    TGAAAGATAC GATTGTGAAC AACTTTCCAT CTCGTGCCGT TAAGCAAGAA
751    TAGTAAATAA TTAGTGTGCA TAACACACGA GGAGGAACAT GAAC ATG GAA
                                                            M   E
801    CAA GGT ACA GTA AAA TGG TTT AAC GCA GAA AAA GGA TTT GGT TTT ATC GA
        Q   G   T   V   K   W   F   N   A   E   K   G   F   G   F   I   E
851    A CGC GAA AAC GGT GAC GAT GTA TTC GTA CAT TTC AGC GCT ATC CAA GGC G
```
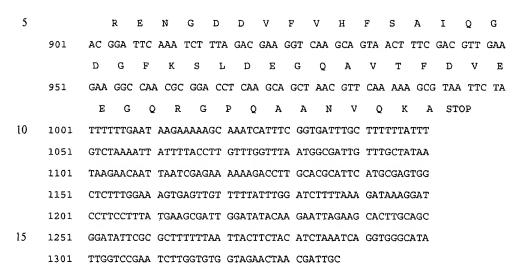
5

```
            R   E   N   G   D   D   V   F   V   H   F   S   A   I   Q   G
   901   AC GGA TTC AAA TCT TTA GAC GAA GGT CAA GCA GTA ACT TTC GAC GTT GAA
            D   G   F   K   S   L   D   E   G   Q   A   V   T   F   D   V   E
   951   GAA GGC CAA CGC GGA CCT CAA GCA GCT AAC GTT CAA AAA GCG TAA TTC TA
            E   G   Q   R   G   P   Q   A   A   N   V   Q   K   A   STOP
```

10
```
  1001   TTTTTTGAAT AAGAAAAAGC AAATCATTTC GGTGATTTGC TTTTTTATTT
  1051   GTCTAAAATT ATTTTACCTT GTTTGGTTTA ATGGCGATTG TTTGCTATAA
  1101   TAAGAACAAT TAATCGAGAA AAAAGACCTT GCACGCATTC ATGCGAGTGG
  1151   CTCTTTGGAA AGTGAGTTGT TTTTATTTGG ATCTTTTAAA GATAAAGGAT
  1201   CCTTCCTTTA TGAAGCGATT GGATATACAA GAATTAGAAG CACTTGCAGC
```

15
```
  1251   GGATATTCGC GCTTTTTTAA TTACTTCTAC ATCTAAATCA GGTGGGCATA
  1301   TTGGTCCGAA TCTTGGTGTG GTAGAACTAA CGATTGC
```